# Intelligent Transportation Systems Joint Program Office (ITS JPO) ITS DataHub

## Data Management Plan

**Draft Report — October 20, 2020**

U.S. Department of Transportation
Office of the Assistant Secretary for Research and Technology
Intelligent Transportation Systems Joint Program Office

Produced by Booz Allen Hamilton
U.S. Department of Transportation
Intelligent Transportation Systems (ITS) Joint Program Office

# Notice

# Revision History

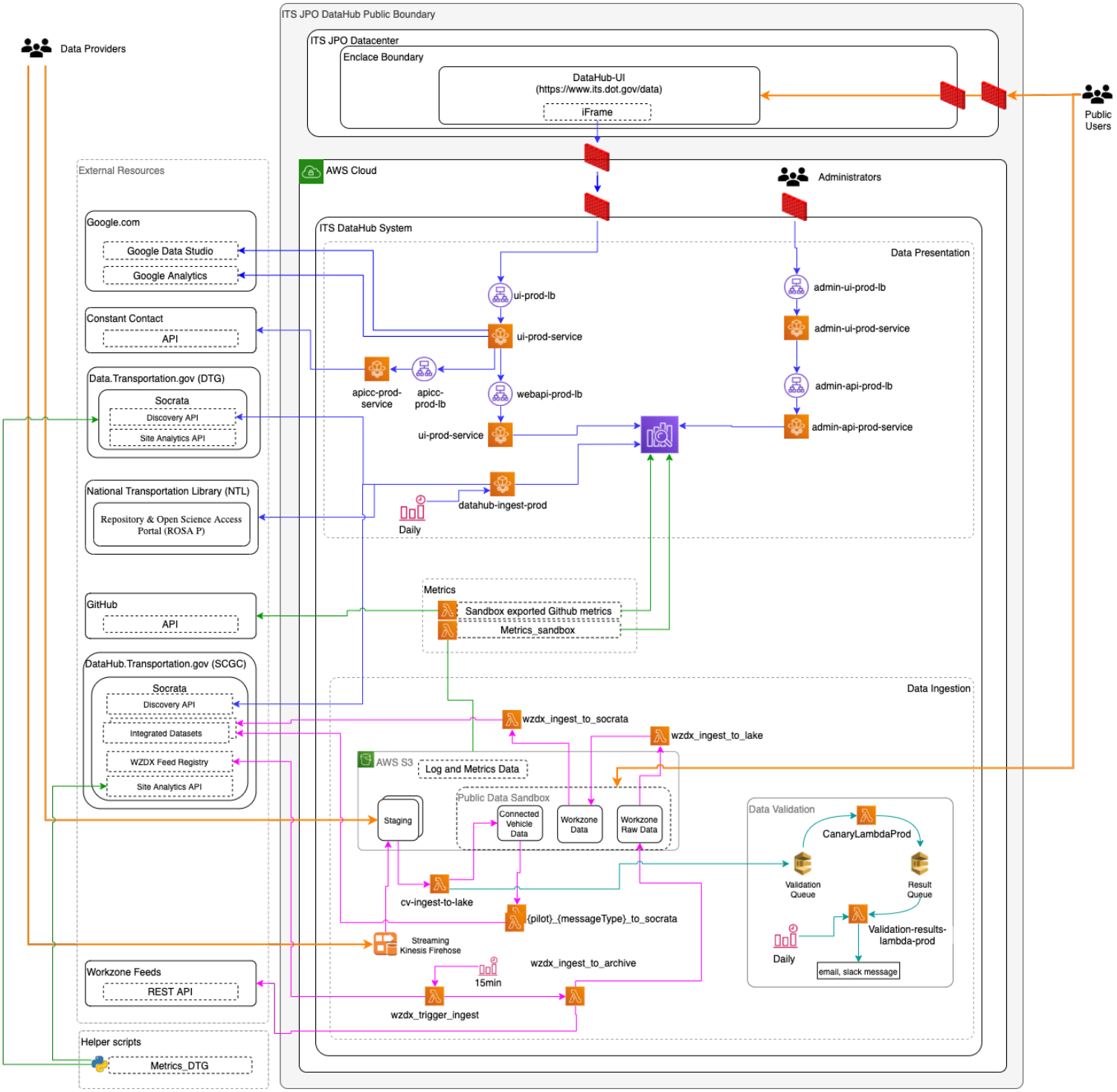
| Date | Description | Version | Author |
|---|---|---|---|
| 08/23/2019 | Document Creation | 1.0 | Booz Allen Team |
| 10/20/2020 | Updated high level data flow diagram, list of relevant documents, data stewardship table, data storage and retention section, and more. Added work zone data archive datasets and modified relevant sections accordingly. Added information relating to ITS DataHub Elasticsearch DB and the datasets it contains. Added datahub.transportation.gov, Google Analytics, Google Data Studio, and Bit.ly related data storage systems and datasets. Updated licenses to CC0 where appropriate. Added Project Performance Measurements. Updated Federal Sponsors of ITS DataHub from Ariel Gold to Mohammad Banihashemi. | 1.1 | Booz Allen Team |

# Table of Contents

## 1. Project Overview

| Project Title | ITS DataHub |
|---|---|
| **Project Goals and Objectives** | The Intelligent Transportation Systems (ITS) Joint Program Office (JPO) and its multimodal partners are dedicated to providing open access to publicly funded research data. By providing access to these data, the United States Department of Transportation (U.S. DOT) aims to:<br><br>• Facilitate early user input into U.S. DOT research efforts, informing multi-phased research projects in real time to improve future deliverables and research goals<br>• Enable third-party research into the effectiveness of these emerging ITS technologies, secondary analysis on primary ITS data collection, and harmonization of data across similar collections<br>• Enable preliminary development of third-party applications using the data generated by ITS JPO projects |
| **Project Description** | ITS DataHub provides a single point of entry to discover U.S. DOT's publicly available ITS research data, including connected vehicle data. The system utilizes shared services, such as the National Transportation Library (NTL), data.transportation.gov (DTG), and datahub.transportation.gov (SCGC) to provide access to timely, discoverable, well-curated research data for public access. ITS DataHub enables a streamlined, consistent, replicable process for data providers to load metadata and data into the system. Data accessible through ITS DataHub is quality-checked, well-documented, and freely available to the public. |
| **Project Lifecycle Phase** | Post-Award |
| **Project Performance Measurements** | Project performance measures include: number of citations/references to datasets in ITS DataHub or SDC, dataset usage (downloads, visualizations, etc), ITS DataHub data page clicks, and establishing baseline metrics on time between planned initial storage date as stated in the preliminary Data Management Plan and data becoming accessible. See the Portfolio-level Metriccs and Targets page for more details. |

## 1.1 High Level Data Flow Diagram



## 1.2 Change Control

The DMP will be reviewed every 6 months and updated as needed to ensure that it continues to reflect the project's current process for data management. When changes are required, the version of the report will be updated, and any modifications will be described at the beginning of the report under a Revision History section.

## 1.3 Relevant Documents

- ITS DataHub Webpage (https://www.its.dot.gov/data/)
- ITS DataHub Design Document (https://usdotjpoode.atlassian.net/wiki/download/attachments/711295786/ITS%20DataHub%20Updated%20Design%20Document_final.docx?api=v2) - last updated August 16, 2019
- ITS DataHub GitHub Page (https://github.com/U.S. DOT-its-jpo-data-portal)
- ITS DataHub Data Retention Schedule (https://usdotjpoode.atlassian.net/wiki/download/attachments/711295634/ITS_DataHub_Disposition_Schedule.xlsx?api=v2) - last updated July 7, 2020
- ITS DataHub System Security Plan (https://usdotjpoode.atlassian.net/wiki/spaces/RDA/pages/592707604/ATO) - last updated July 24, 2019

## 2. Data Overview

| ID | Dataset Title | Description | Type / Scale | Collection Method | Data File Format(s) | Access Tools |
|---|---|---|---|---|---|---|
| N/A | ITS DataHub Configuration and Data Asset Inventory | This data consists of the metadata of all ITS JPO data assets that have been cataloged with ITS DataHub. This data also includes configurations for the ITS DataHub website, including but not limited to, the configurations for “projects”, “data types”, and “engagement pop ups”. | Metadata of ITS DataHub and the data assets discoverable via ITS DataHub. | The metadata of the dataset is added into the appropriate data-storage system (NTL, DTG, SCGC). The ITS DataHub ingestion process then hits the catalog APIs at NTL, DTG, and SCGC daily to retrieve and save the most up-to-date dataset metadata on those platforms.<br><br>Configurations of ITS DataHub are added through the Admin User Interface (UI) of ITS DataHub. | Data is stored in Elasticsearch Database but can be easily exported as .json files. | Text editor |

| ID | Dataset Title | Description | Type / Scale | Collection Method | Data File Format(s) | Access Tools |
|---|---|---|---|---|---|---|
| N/A | ITS Sandbox Metrics | This data consists of usage metrics of Connected Vehicle Pilot (CVP) sandbox as well as the usage metrics of the sandbox_exporter GitHub repository. | Numerical metrics of CVP sandbox and one GitHub repository. | Object-level logging to AWS CloudTrail is enabled for the CVP sandbox S3 bucket, and ITS DataHub runs a script daily to aggregate those logs as well as retrieve metrics from GitHub's API. | Data is stored in Elasticsearch Database but can be easily exported as .json files. | Text editor |
| N/A | Google Analytics Metrics | This data consists of usage metrics of ITS DataHub and ITS sandbox web interface. | Numerical metrics on web user interactions with ITS DataHub and the ITS sandbox web interface. | Data collected through Google Analytics and Google Tag Manager. | Data is stored in Google Analytics but can be easily exported as .csv and Excel files. | Browser or Excel. |
| N/A | Bit.ly Metrics | The data consists of metrics related to bit.ly links we have generated. | Numerical metrics on web user interactions with our bit.ly links. | Metrics related to interactions with our Bit.ly links are collected by Bit.ly. When evaluating effectiveness of promotion events, metrics for related Bit.ly URLs are viewed via Bit.ly's website and recorded in our Metrics report on Confluence. | Data is stored within Bit.ly and on the Metrics page on Confluence page. The Metrics page on Confluence can be exported as PDF. | Browser or PDF viewer. |

| ID | Dataset Title | Description | Type / Scale | Collection Method | Data File Format(s) | Access Tools |
|---|---|---|---|---|---|---|
| N/A | ITS DataHub Data Asset Usage Metrics | This data consists of usage metrics of data assets that have been cataloged with ITS DataHub. | Numerical data. | Exports of NTL metrics are sent to ITS DataHub for ingestion monthly. ITS DataHub also runs a script daily to collect metrics from DTG. | Data is stored in Google Drive but can be easily exported as .csv and Excel files. | Browser or Excel. |
| N/A | Wyoming Department of Transportation (WYDOT) Connected Vehicle (CV) Pilot Deployment Sandbox Data | This data consists of Basic Safety Messages (BSMs) and Traveler Information Messages (TIMs) generated by participant and public transportation vehicles onboard units (OBU) and transmitted to road-side units (RSU) located throughout the Project Study area. | Numerical data, text sequences, positional data (e.g. latitude and longitude). | Experimental with sensors placed throughout the test area and on the car collecting daily information. | Newline json files | Text editor |

| ID | Dataset Title | Description | Type / Scale | Collection Method | Data File Format(s) | Access Tools |
|---|---|---|---|---|---|---|
| N/A | Tampa Hillsborough Expressway Authority (THEA) CV Pilot Deployment Sandbox Data | This data consists of Basic Safety Messages (BSMs), Traveler Information Messages (TIMs), and Signal Phasing and Timing (SPaT) messages generated by participant and public transportation vehicles onboard units (OBU) and transmitted to road-side units (RSU) located throughout the Project Study area. | Numerical data, text sequences, positional data (e.g. latitude and longitude). | Experimental with sensors placed throughout the test area and on the car collecting daily information. | Newline json files | Text editor |
| N/A | ITS Data Sandbox: Work Zone Raw Data Archive | This data consists of snapshots of data retrieved from active data feeds listed in the Work Zone Feed Registry dataset, following the Work Zone Data Exchange (WZDx) specifications. | Numerical data, text sequences, positional data (e.g. latitude and longitude). | A HTTP GET request is sent to each active feed listed in the Work Zone Feed Registry in regular intervals based on each feed's update frequency. Each response is then saved in a text file without additional processing and uploaded to the appropriate directory in the sandbox S3 bucket. | Text files | Text editor |

| ID | Dataset Title | Description | Type / Scale | Collection Method | Data File Format(s) | Access Tools |
|---|---|---|---|---|---|---|
| N/A | ITS Data Sandbox: Work Zone Semi-Processed Data Archive | This data consists of semi-processed data retrieved from active data feeds listed in the Work Zone Feed Registry dataset, following the Work Zone Data Exchange (WZDx) specifications. Each file is a newline JSON file and contains the work zone statuses for a particular work zone for the month. | Numerical data, text sequences, positional data (e.g. latitude and longitude). | Each file saved into the Work Zone Raw Data Archive is processed and each work zone status is appended to the appropriate file in the sandbox S3 bucket if it is different from the last saved work zone status for the work zone. The first status and the most recent status retrieved for the month are always archived, as are statuses flanking any detected change. At least one status is also archived for each day even if the status of the work zone has not changed. | Newline json files | Text editor |

| ID | Dataset Title | Description | Type / Scale | Collection Method | Data File Format(s) | Access Tools |
|---|---|---|---|---|---|---|
| N/A | (Planned) New York City Department of Transportation CV Pilot Deployment Sandbox Data | - | Numerical data, text sequences, positional data (e.g. latitude and longitude) | Experimental with sensors placed throughout the test area and on the car collecting daily information. | Newline json files | Text editor |

# 3. Data Stewardship

## 3.1 Data Owner and Steward

| Dataset Title | Data Owner | Data Steward | Federal Sponsor |
|---|---|---|---|
| ITS DataHub Configuration and Data Asset Inventory | U.S. DOT | ITS JPO | Mohammad Banihashemi |
| ITS Sandbox Metrics | U.S. DOT | ITS JPO | Mohammad Banihashemi |
| Google Analytics Metrics | U.S. DOT | ITS JPO | Mohammad Banihashemi |
| Bit.ly Metrics | U.S. DOT | ITS JPO | Mohammad Banihashemi |
| ITS DataHub Data Asset Usage Metrics | U.S. DOT | ITS JPO | Mohammad Banihashemi |
| Wyoming Department of Transportation (WYDOT) Connected Vehicle (CV) Pilot Deployment Sandbox Data | U.S. DOT | Tony English | Walter During |
| Tampa Hillsborough Expressway Authority (THEA) CV Pilot Deployment Sandbox Data | U.S. DOT | Sisinnio Concas | Govindarajan Vadakpat |
| ITS Data Sandbox: Work Zone Raw Data Archive | U.S. DOT | ITS JPO | Todd Peterson |
| ITS Data Sandbox: Work Zone Semi-Processed Data Archive | U.S. DOT | ITS JPO | Todd Peterson |
| (Planned) New York City Department of Transportation CV Pilot Deployment Sandbox Data | Jointly owned by NYDOT and NYC CVPD participating organizations | TBD | Kate Hartman |

## 3.2 Access Level

### 3.2.1 Public Access Level

Can all data from this project be shared with the public or is controlled-access required for at least some of the data?

☒ All Public Access ☐ Some/All Controlled-Access

### 3.2.2 Access Requests

ITS DataHub is intended to be a publicly open source of data. In most cases, all data included in ITS DataHub is made available for public viewing. Further explanations of data access policies are below.

**ITS DataHub Configuration and Data Asset Inventory**

ITS DataHub Configuration and Data Asset Inventory is not fully publicly available. The dataset includes fields used for internal tracking purposes as well as public-facing fields for configuring and populating the ITS DataHub web application. The public-facing fields in the dataset are publicly available and can be searched through the ITS DataHub portal at https://www.its.dot.gov/data/. No access requests approval process is required.

**ITS Sandbox Metrics**

A portion of ITS Sandbox Metrics is readily available via the ITS DataHub Metrics Dashboard at https://its.dot.gov/data/#/metrics without requiring additional access requests approval. To access the full sandbox metrics, users may contact ITS DataHub Support Staff at data.itsjpo@dot.gov to request for a data dump.

**Google Analytics Metrics**

The Sandbox Web Interface pageview portion of Google Analytics is readily available via the ITS DataHub Metrics Dashboard at https://its.dot.gov/data/#/metrics without requiring additional access requests approval. To access the full dataset, users may contact ITS DataHub Support Staff at data.itsjpo@dot.gov to request for a data dump.

**Bit.ly Metrics**

The Bit.ly metrics are available via internal Metrics report on Confluence. To access the full report, users may contact ITS DataHub Support Staff at data.itsjpo@dot.gov to request for an export.

**ITS DataHub Data Asset Usage Metrics**

All of the data asset usage metrics are readily available via the ITS DataHub Metrics Dashboard at https://its.dot.gov/data/#/metrics without requiring additional access requests approval.

**ITS Sandbox Data**

ITS Sandbox data are stored in Amazon Web Services (AWS) Simple Storage Services (S3) buckets. Currently, ITS Sandbox data includes data from Connected Vehicle Pilots and data from Work Zone Data Exchange Pilots.

*For Data Consumers:*

All data stored in public/published buckets is available to be viewed and downloaded by anyone. Data consumers are not able to access ingest buckets which holds staging data. This access control is enforced by the ITS DataHub Support Staff via AWS role enforcement.

*For Data Providers of CV Pilot Deployment Data:*

When a dataset is accepted for storage in the ITS Sandbox through scheduled batch uploads, the data provider is provided with two ingestion buckets: one for production and one for testing. These ingestion buckets allow data providers to deposit data to ITS DataHub while allowing ITS DataHub to remain the only publisher of the public bucket. Data providers are given read and write access to the ingestion bucket specific to the dataset(s) they contribute to. This access control is enforced by the ITS DataHub Support Staff via AWS role enforcement.

*For Data Providers of Work Zone Data:*

Since ITS DataHub ingests Work Zone data by hitting the active work zone data feeds hosted by the data provider, no additional AWS privileges are given to data providers of work zone data.

*For ITS DataHub Support Staff:*

The ITS DataHub Support Staff has both read and write access to all S3 buckets in the ITS Sandbox. The ITS DataHub Support Staff also manages the provisioning of accounts with S3 access.

### 3.2.3 Relevant Privacy and/or Security Agreements

This section is not applicable since the data assets are public. However, ITS DataHub does have a memorandum of understanding with NTL.

## 3.3 Re-Use, Redistribution, and Derivative Products Policies

*This section is required for* all *anticipated datasets of the project.*

| Dataset Title | License Used | Reason(s) for Non-Open License |
|---|---|---|
| ITS DataHub Configuration and Data Asset Inventory | CC0 1.0 Universal Public Domain Dedication | N/A |
| ITS Sandbox Metrics | CC0 1.0 Universal Public Domain Dedication | N/A |
| Google Analytics Metrics (ITS DataHub Dataset) | CC0 1.0 Universal Public Domain Dedication | N/A |
| Bit.ly Metrics (ITS DataHub Dataset) | CC0 1.0 Universal Public Domain Dedication | N/A |
| ITS DataHub Data Asset Usage Metrics | CC0 1.0 Universal Public Domain Dedication | N/A |
| Wyoming Department of Transportation (WYDOT) Connected Vehicle (CV) Pilot Deployment Sandbox Data | CC BY-SA 4.0 Attribution-ShareAlike 4.0 International | N/A |
| Tampa Hillsborough Expressway Authority (THEA) CV Pilot Deployment Sandbox Data | CC BY-SA 4.0 Attribution-ShareAlike 4.0 International | N/A |
| ITS Data Sandbox: Work Zone Raw Data Archive | CC0 1.0 Universal Public Domain Dedication | N/A |
| ITS Data Sandbox: Work Zone Semi-Processed Data Archive | CC0 1.0 Universal Public Domain Dedication | N/A |
| (Planned) New York City Department of Transportation CV Pilot Deployment Sandbox Data | CC BY-SA 4.0 Attribution-ShareAlike 4.0 International | N/A |

## 3.4 Data Storage and Retention

### 3.4.1 Storage Systems

| Data Storage System Name | Data Storage System Type | Dataset Title(s) | Initial Storage Date | Frequency of Update | Archiving and Preservation Period |
|---|---|---|---|---|---|
| ITS DataHub Elasticsearch | U.S. DOT-managed – Public System | ITS DataHub Configuration and Data Asset Inventory<br><br>ITS Sandbox Metrics | N/A | Daily | This data will persist as long as the ITS DataHub web application requires the data to operate. |
| Google Analytics | U.S. DOT-managed account | Google Analytics Metrics | N/A | Daily | This data will persist as long as the ITS DataHub team requires the data to operate. |
| Bit.ly | U.S. DOT-managed account | Bitly Metrics | N/A | Daily | This data will persist as long as the ITS DataHub team requires the data to operate. |
| Google Drive | U.S. DOT-managed account | ITS DataHub Data Asset Usage Metrics | N/A | Irregularly, as NTL metrics are received, generally monthly | This data will persist as long as the ITS DataHub team requires the data to operate. |
| Data.transportation.gov (DTG) | U.S. DOT-managed – Public System | ITS DataHub Configuration and Data Asset Inventory | See ITS DataHub Data Retention Schedule in Relevant Documents section | Irregularly, as new datasets become available | Five years or until no longer needed for business purposes, whichever is later.<br><br>See ITS DataHub Data Retention Schedule in Relevant Documents section. |
| Datahub.transportation.gov (SCGC) | U.S. DOT-managed – Public System | ITS DataHub Configuration and Data Asset Inventory | See ITS DataHub Data Retention Schedule in Relevant Documents section | Irregularly, as new datasets become available | Five years or until no longer needed for business purposes, whichever is later.<br><br>See ITS DataHub Data Retention Schedule in Relevant Documents section. |
| National Transportation Library (NTL) | U.S. DOT-managed – Public System | ITS DataHub Configuration and Data Asset Inventory | See ITS DataHub Data Retention Schedule in Relevant Documents section | Irregularly, as new datasets become available | Five years or until no longer needed for business purposes, whichever is later.<br><br>See ITS DataHub Data Retention Schedule in Relevant Documents section. |

| Data Storage System Name | Data Storage System Type | Dataset Title(s) | Initial Storage Date | Frequency of Update | Archiving and Preservation Period |
|---|---|---|---|---|---|
| ITS Data Sandbox: Connected Vehicle Pilot Data Archive | U.S. DOT-managed – Public System | Wyoming Department of Transportation (WYDOT) Connected Vehicle (CV) Pilot Deployment Sandbox Data<br><br>Tampa Hillsborough Expressway Authority (THEA) CV Pilot Deployment Sandbox Data<br><br>(Planned) New York City Department of Transportation CV Pilot Deployment Sandbox Data | See ITS DataHub Data Retention Schedule in Relevant Documents section | Daily | Five years or until no longer needed for business purposes, whichever is later.<br><br>See ITS DataHub Data Retention Schedule in Relevant Documents section. |
| ITS Data Sandbox: Work Zone Data Archive | U.S. DOT-managed – Public System | ITS Data Sandbox: Work Zone Raw Data Archive<br><br>ITS Data Sandbox: Work Zone Semi-Processed Data Archive | See ITS DataHub Data Retention Schedule in Relevant Documents section | Irregularly and up to once every 30 minute, depending on the update frequency of each feed. | Two years since when the data feed is first added to the Work Zone Data Exchange (WZDx) Feed Registry or no longer needed for business purposes.<br><br>See ITS DataHub Data Retention Schedule in Relevant Documents section. |

### 3.4.2 Data Storage System Description

**ITS DataHub Elasticsearch Database:**

A U.S. DOT-managed data storage system that stores non-sensitive configuration information of ITS DataHub, a subset of metadata of datasets discoverable through ITS DataHub, and metrics information on ITS Sandbox and related GitHub repository. This data storage system runs on AWS Elasticsearch.

**Google Analytics:**

A U.S. DOT-managed account on Google Analytics that stores Google Analytics data on ITS DataHub and ITS Sandbox web interface.

**Bit.ly:**

A U.S. DOT-managed account on Bit.ly that stores metrics data on Bit.ly links generated for ITS DataHub.

**Google Drive:**

A U.S. DOT-managed account on Google Drive that stores metrics data of datasets discoverable through ITS DataHub.

**Data.transportation.gov (DTG):**

A U.S. DOT-managed public-access data storage system that stores non-sensitive, primary research data and data information such as metadata and data dictionaries. Primary research data is data collected during research that has not had significant analysis performed on it. Data.transportaiton.gov provides federation to datasets hosted in datahub.transportation.gov and allows community users to create derived assets such as visualizations and filters, from datasets hosted on DTG. This data storage system is Socrata-based.

**Datahub.transportation.gov (SCGC):**

A U.S. DOT-managed public-access data storage system that stores non-sensitive, primary research data and data information such as metadata and data dictionaries. Primary research data is data collected during research that has not had significant analysis performed on it. Datahub.transportation.gov is the newer version of Data.transportation.gov and has additional functionality for non-community users. This data storage system is Socrata Connected Government Cloud based.

**National Transportation Library (NTL):**

A U.S. DOT-managed public-access data storage system that stores non-sensitive, derived research data. Derived research data is data derived thorough analysis of other data and shared as part of a final report to communicate the methodology and findings of a research project. ITS JPO derived research data is stored in NTL's ROSA P catalog. Derived datasets tend to be smaller than primary datasets. These datasets should ensure they follow the **public access instructions provided by NTL**, which the ITS JPO references in its materials. This data storage system is ROSA P-based.

**ITS Data Sandbox:**

A U.S. DOT-managed public-access data storage system that stores non-sensitive, primary research data that is not suitable to be fully housed in DTG usually due to its large or continuously expanding size, immaturity, and/or non-standard or non-finalized format and structure. Data providers whose data id submitted to the ITS Sandbox must work with the ITS JPO to determine the best way to deposit data into the ITS Sandbox. Usually this involves set up of an automated data ingestion pipeline. Once datasets in the ITS Sandbox reach maturity, they may be moved to another system, usually **DTG**. In some cases, the data ingestion processes may continue to ingest data into the ITS Sandbox even once these datasets move to another system. This data storage system is AWS S3-based.

### 3.4.3 Cybersecurity Policies

Cybersecurity policies for data stored in NTL, DTG, SCGC, Google Analytics, Bit.ly, and Google Drive are handled by each platform using their resources. For data that is in AWS, the cybersecurity policies are handled through IAM user configurations.

### 3.4.4 Data Security Policies and Procedures

Security implementations for data stored in NTL, DTG, SCGC, Google Analytics, Bit.ly, and Google Drive are handled by each platform using their resources. For ITS Data Sandbox, data files are stored in S3 buckets in the AWS account and are unencrypted at rest. ITS DataHub Elasticsearch Database data are also unencrypted.

### 3.4.5 Back-up and Recovery Policies and Procedures

The majority of the data that comes through ITS DataHub is stored on external systems (e.g. NTL, DTG, SCGC, Google Analytics, Bit.ly, Google Drive) that the ITS DataHub Support Staff does not directly control. To further ensure backup availability, backup buckets will be placed in different AWS Region/Availability Zones.

ITS DataHub Configuration and Data Asset Inventory as well as the ITS JPO Metrics data in the ITS DataHub Elasticsearch Database (AWS Elasticsearch-based) is backed up daily through cluster-level snapshots with no specified retention limits.

# 4. Data Standards

## 4.1 Collection Format Standards

| Dataset Title | Data Collection Format Standard(s) | Data Collection Format Standard Uniform Research Identifiers(s) (URIs) | Open or Proprietary? | Collection Format Rationale |
|---|---|---|---|---|
| ITS DataHub Configuration and Data Asset Inventory | Not following a particular standard | N/A | Open | N/A |
| ITS Sandbox Metrics | Not following a particular standard | N/A | Open | N/A |
| Google Analytics Metrics | Not following a particular standard | N/A | Open | N/A |
| Bit.ly Metrics | Not following a particular standard | N/A | Open | N/A |
| ITS DataHub Data Asset Usage Metrics | Not following a particular standard | N/A | Open | N/A |

| Dataset Title | Data Collection Format Standard(s) | Data Collection Format Standard Uniform Research Identifiers(s) (URIs) | Open or Proprietary? | Collection Format Rationale |
|---|---|---|---|---|
| Wyoming Department of Transportation (WYDOT) Connected Vehicle (CV) Pilot Deployment Sandbox Data | Operational Data Environment (ODE) output schema | N/A | Open | Using the ITS ODE within intelligent transportation deployments increases data fluidity and interoperability while meeting operational needs and protecting user privacy. ITS ODE takes data following the SAE J2735 standard and transforms them to non-proprietary open formats to be consumed by client applications. |
| Tampa Hillsborough Expressway Authority (THEA) CV Pilot Deployment Sandbox Data | SAE J2735 | https://saemobilus.sae.org/content/j2735_201603 | Proprietary | This Standard is the fifth edition of the message set dictionary. The changes made from prior editions include revising the content to reflect a uniform use of unaligned packed coding rules, a common message framework, the further refinement of several existing messages due to deployment experience, and the addition of a preliminary Personal Safety Message for use with vulnerable road users. |
| ITS Sandbox on AWS S3: Work Zone Data Archives | Work Zone Data Exchange (WZDx) Specifications | N/A | Open | This specification enables infrastructure owners and operators (IOOs) to make harmonized work zone data available for third-party use and is developed in the open through USDOT Federal Highway Administration (FHWA) and ITS JPO in collaboration with data providers and data users. |
| (Planned) New York City Department of Transportation CV Pilot Deployment Sandbox Data | - | - | - | - |

## 4.2 Versioning

### ITS DataHub Configuration and Data Asset Inventory

Versioning is not applicable for ITS DataHub Configuration and Data Asset Inventory as it gets updated as new data assets get cataloged. Any data dump of the dataset that is exported from the databases and shared will have the retrieved date and timestamp in its filename as version control.

### ITS Sandbox Metrics

Versioning is not applicable for ITS Sandbox Metrics as it is a dataset that is being updated daily. Any data dump of the ITS Sandbox Metrics that is exported from the database and shared will have the retrieved date and timestamp in its filename as version control.

### Google Analytics Metrics

Versioning is not applicable for Google Analytics Metrics as it is a dataset that is being updated continuously. Any data dump of the Google Analytics Metrics that is exported from the database and shared will have the retrieved date and timestamp in its filename as version control.

### Bit.ly Metrics

Versioning is not applicable for Bit.ly Metrics as it is a dataset that is being updated Scontinuously. Any export of the Bit.ly Metrics will have the retrieved date and timestamp in its filename as version control.

### ITS DataHub Data Asset Usage Metrics

Versioning is not applicable for ITS DataHub Data Asset Usage Metrics as it is a dataset that is being updated daily. Any data dump of the ITS DataHub Data Asset Usage Metrics that is exported from the database and shared will have the retrieved date and timestamp in its filename as version control.

### ITS Data Sandbox: Connected Vehicle Pilot

Versioning of data in the Connected Vehicle Pilot Data Sandbox is tracked through the “schemaVersion” field in the “metadata” field of each record (e.g. record[“metadata”][“schemaVersion”).

### ITS Data Sandbox: Work Zone Data Archive

Versioning of data specification in the Work Zone Raw Data Archive is tracked through the version field in each feed snapshot, which is in the `record["WZDx"]["Header"]["versionNo"]` field for version 1.1 feeds and in the `record["road_event_feed_info"]["version"]` field for version 2 feeds. Versioning of data specification in the Work Zone Semi-Processed Data Archive is tracked through the file name, which follows this format:
`{identifier}_{beginLocation_roadDirection}_{year}{month}_v{version}.txt`.

## 4.3 Metadata and Data Dictionary

| Dataset Title | Metadata Standards Used | Metadata Discoverable (Y/N) | Data Dictionary Discoverable (Y/N) | Metadata and Data Dictionary Access |
|---|---|---|---|---|
| ITS DataHub Configuration and Data Asset Inventory | N/A | Y | Y | https://github.com/usdot-its-jpo-data-portal/its-dmp-data-dictionary/blob/master/its_jpo_dataset_inventory_data_dictionary.csv |
| ITS Sandbox Metrics | N/A | Y | Y | https://github.com/usdot-its-jpo-data-portal/its-dmp-data-dictionary/blob/master/its_jpo_metrics_data_dictionary.csv |
| Google Analytics Metrics | N/A | Y | Y | The full set of metrics available can be found at https://developers.google.com/analytics/devguides/collection |
| Bit.ly Metrics | N/A | Y | Y | The full set of metrics available can be found at https://dev.bitly.com/docs/tutorials/retrieve-metrics |
| ITS DataHub Data Asset Usage Metrics | N/A | Y | Y | https://github.com/usdot-its-jpo-data-portal/its-dmp-data-dictionary/blob/master/its_datahub_data_asset_usage_metrics_data_dictionary.csv |
| Wyoming Department of Transportation (WYDOT) Connected Vehicle (CV) Pilot Deployment Sandbox Data | ODE output schema | Y | Y | https://github.com/U.S._DOT-jpo-ode/jpo-ode/blob/master/docs/ODE_Output_Schema_Reference.docx |

| Dataset Title | Metadata Standards Used | Metadata Discoverable (Y/N) | Data Dictionary Discoverable (Y/N) | Metadata and Data Dictionary Access |
|---|---|---|---|---|
| Tampa Hillsborough Expressway Authority (THEA) CV Pilot Deployment Sandbox Data | SAE J2735 and J2945/1 standards | Y | Y | https://github.com/usdot-its-jpo-data-portal/its-dmp-data-dictionary/blob/master/thea_bsm_sandbox_data_dictionary.csv<br><br>https://github.com/usdot-its-jpo-data-portal/its-dmp-data-dictionary/blob/master/thea_spat_sandbox_data_dictionary.csv<br><br>https://github.com/usdot-its-jpo-data-portal/its-dmp-data-dictionary/blob/master/thea_tim_sandbox_data_dictionary.csv |
| ITS Data Sandbox: Workzone Raw Data Archive | WZDx Specification (v1.1, v2) | Y | Y | https://github.com/usdot-jpo-ode/jpo-wzdx/tree/v1.1<br><br>https://github.com/usdot-jpo-ode/jpo-wzdx/tree/v2.0 |
| ITS Data Sandbox: Workzone Semi-Processed Data Archive | WZDx Specification (v1.1, v2) | Y | Y | https://github.com/usdot-jpo-ode/jpo-wzdx/tree/v1.1<br><br>https://github.com/usdot-jpo-ode/jpo-wzdx/tree/v2.0 |
| (Planned) New York City Department of Transportation CV Pilot Deployment Sandbox Data | - | Y | Y | - |

### 4.3.1 Metadata Description

All metadata are discoverable through this Data Management Plan, which will be publicly available through National Transportation Library. All data dictionaries, if not available with the data storage system's documentations, will be hosted on the "its-dmp-data-dictionary" GitHub repository (https://github.com/usdot-its-jpo-data-portal/its-dmp-data-dictionary). The data dictionary will also be linked in the documentation of associated GitHub repository whenever applicable. For example, the data dictionary of the Wyoming, Tampa, and NYC Connected Vehicle Pilot will be

hosted in the "dmp-data-dictionary" GitHub repository but referenced in the ITS Sandbox AWS S3 Explorer page (http://U.S. DOT-its-cvpilot-public-data.s3.amazonaws.com/index.html) and other associated GitHub repositories.

## 5. Glossary of Terms

**Amazon Web Services (AWS) Relational Database Service (RDS):** Relational database as a service on Amazon Web Services (AWS).

**Bit.ly:** An URL shortening service and a link management platform that allow users to track engagement with the shortened links.

**Data Dictionary:** A machine readable file that contains the most up to date version of metadata which describes and defines a dataset's elements, fields, or column headings. This metadata describes the meaning and purposes of each data element, provides a list of valid values for each variable, and helps one maintain consistency in how the data is described and categorized.

**Data.transportation.gov (DTG):** A U.S. DOT-managed public-access data storage system that stores non-sensitive, primary research data and data information such as metadata and data dictionaries. Primary research data is data collected during research that has not had significant analysis performed on it.

**Datahub.transportation.gov (SCGC):** A U.S. DOT-managed public-access data storage system that stores non-sensitive, primary research data and data information such as metadata and data dictionaries. Primary research data is data collected during research that has not had significant analysis performed on it. Datahub.transportation.gov is the newer version of Data.transportation.gov and has additional functionality for non-community users. This data storage system is Socrata Connected Government Cloud based.

**Derived Research Data:** Derived research data is data derived thorough analysis of other data and shared as part of a final report to communicate the methodology and findings of a research project. Derived datasets tend to be smaller than primary datasets.

**ITS DataHub:** Website maintained by the ITS JPO that provides a single point of entry to discover available U.S. DOT ITS research data. Currently accessible at https://its.dot.gov/data/.

**ITS Operational Data Environment (ODE):** The ITS ODE is a real-time virtual data router that ingests and processes operational data from various connected devices - including vehicles, infrastructure, and traffic management centers - and distributes it to other devices and subscribing transportation management applications. Use of ITS ODE within intelligent transportation deployments increases data fluidity and interoperability while meeting operational needs and protecting user privacy.

**ITS Sandbox:** A U.S. DOT-managed public-access data storage system that stores non-sensitive, primary research data that is not suitable to be fully housed in DTG usually due to its extreme size, immaturity, and/or non-standard or non-finalized format and structure.

**National Transportation Library (NTL):** A U.S. DOT-managed public-access data storage system that stores non-sensitive, derived research data. This data storage system is powered by ROSA P catalog.

U.S. Department of Transportation
ITS Joint Program Office-HOIT
1200 New Jersey Avenue, SE
Washington, DC 20590

Toll-Free “Help Line” 866-367-7487
www.its.dot.gov